

作业二：探索 iris(鸢尾花)数据集

```
>>> from sklearn import datasets
>>> import numpy as np

>>> iris = datasets.load_iris()
```

一、探索 iris 的以下属性：

- .DESCR:数据集总体描述
- .data:所有观测实例的特征数据(numpy.ndarray 类型)
- .data.shape:维信息
- .feature_names:特征名(sepal 萼片,petal 花瓣)
- .target:所有实例的类别值(numpy.ndarray 类型),0/1/2
- .target.shape:维信息
- .target_names:类别名称(setosa/versicolor/virginica)

例如,

```
>>> print iris.data
>>> print iris.target
```

二、探索各特征的最小值,最大值,均值,中位数,标准差.

三、探索各特征之间,以及特征与目标之间的相关性(相关系数).

例如,

```
>>> np.corrcoef(iris.data.T[0],iris.target)
array([[1.          , 0.78256123],
       [0.78256123, 1.          ]])
```

说明萼片长度与类别有一定相关度.



setosa



versicolor



virginica

作业上传(限时 2 周):

<ftp://public.sjtu.edu.cn>

qiwu.zhou

public